引用格式：乔晗, 徐君如. 基于LDA模型与政策工具的中国数据主权政策研究. 中国科学院院刊, 2024, 39(3): 498-508, doi: 10.16418/j.issn.1000-3045.20231115001.
Qiao H, Xu J R. Research on Chinese data sovereignty policy based on LDA model and policy instruments. Bulletin of Chinese Academy of Sciences, 2024, 39(3): 498-508, doi: 10.16418/j.issn.1000-3045.20231115001. (in Chinese)

基于LDA模型与政策工具的中国数据主权政策研究

乔 晗* 徐君如

1 中国科学院大学 经济与管理学院 北京 100190

2 中国科学院大学 数字经济监测预测预警与政策仿真教育部哲学社会科学实验室（培育） 北京 100190

摘要 在数字经济发展与总体国家安全观的双重时代背景下，数据主权已成为国家主权的重要组成部分。各主要国家和地区积极开展数据主权战略部署，在数据资源、数据技术和数据规则方面展开激烈竞争与博弈。文章采用政策文本分析方法研究我国数据主权政策，运用LDA（Latent Dirichlet Allocation）主题模型和政策工具量化分析中国数据主权政策的过程演化、主题特征，并综合考虑全球数据主权态势，提出4条政策建议：积极主导和参与国际规则制定；优化数据出境安全评估流程；完善个人信息出境标准合同模版；强化数据安全法治保障。

关键词 数据主权，数据安全，政策工具，LDA模型

DOI 10.16418/j.issn.1000-3045.20231115001

CSTR 32128.14.CASbulletin.20231115001

数据主权是指一个国家对其政权管辖范围内的网络设施、数据主体、数据行为和数据资源及相关数据产品具有生成、传播、管理、控制、利用和保护的主导权[1-3]，其正在成为数字时代国家主权的重要组成部分。美国、欧盟和中国等国家和地区均认识到数据资源所具有的战略价值，积极开展数据主权战略部署。由于技术能力和经济发展状况不同等因素，各国采取了差异化的数据主权政策。

数据主权政策对于维护国家安全、保障国家利益具有重要作用，近年来引起学术界较为广泛的关注。

*通信作者

资助项目：国家自然科学基金重大项目（72192843），国家社会科学基金项目（22VRC055）
修改稿收到日期：2024年3月4日

有的学者从国家主体的视角提出了数据主权概念[4-7]，也有学者从组织和个人主体的视角界定了数据主权内涵[8,9]。冉从敬[10]定性总结了数据主权战略部署的全球态势。郑琳等[11]阐释了国家数据主权概念，并归纳分析了欧美数据主权战略。然而，已有研究主要通过定性归纳方法从国家宏观层面剖析数据主权战略的特点，但缺乏关于数据主权政策文本内涵和特征的定量分析。本文采用政策文本分析方法①研究我国数据主权政策，运用LDA（Latent Dirichlet Allocation）主题模型②和政策工具③定量分析中国数据主权政策的演化、态势，基于此与全球数据主权政策进行国际比较并提出对策与建议。这一研究既拓展了政策文本分析方法的应用领域，将其应用于数据主权领域的政策分析，也为优化设计我国数据主权政策提供方法论指导。

1 数据主权政策研究的分析框架构建与设计

1.1 研究方法与框架维度

本文从政策工具的角度出发，运用LDA主题模型对中国2010—2022年的数据主权相关政策进行分析与量化。本文建立了数据主权政策工具的分析框架（图1）。① 对政策文本进行统计分析，围绕政策时间、政策机构和政策类型分布展开，探究数据主权政策的演化过程、机构分布和政策效力。② 供给型、环境型和需求型政策工具分类法在既有研究中最为常见[12]，本文采用此种政策工具的维度划分对数据主权政策进行编码分析。③ 运用LDA主题模型挖掘数据主权政策文本中的核心主题词，并进行可视化展示。

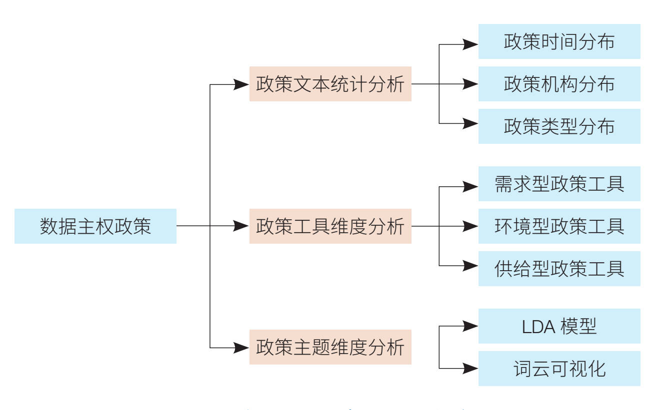


图1 数据主权政策研究框架

Figure 1 Data sovereignty policy research framework

1.2 数据来源及编码

本文选用“北大法宝在线数据库”④作为数据主权相关政策文本的收集来源，以“数据主权”“数据安全”“数据跨境”“国家安全”等关键词搭配组合进行全文检索，筛选“中央法规”政策文件作为分析文本，共计收集45份政策文件（表1展示了部分数据主权政策文本）。对纳入样本的45份中央法规政策文件进行摘录和编码（部分数据主权政策文件内容编码情况见表2），以便更清晰地标注不同政策与具体条款中包含的政策工具，将非结构化的文本数据进行转换用于定量分析。

2 数据主权政策总体情况

2.1 政策时间演化分布

根据政策文本的发布时间统计结果，中国数据主权相关政策始于2012年5月工信部发布的《互联网行业“十二五”发展规划》；“数据主权”这一名词首次

① 政策文本分析是政策研究的主要范式之一，它以政策内涵、政策工具等为主要研究内容，利用定量分析对文本条例作出深入阐释与分析，对于政策解读和实施具有重要意义。

② LDA主题模型是针对海量长文本数据进行概率主题建模的自然语言处理模型，可以抽象为三层贝叶斯模型“文本—主题—词”。作为识别文档主题分布的典型方法，LDA主题模型在文本主题挖掘、文本情感分析等领域具有广泛应用。

③ 已有研究从不同角度对政策工具进行分类，包括供给型、环境型和需求型政策工具的界定方式，市场工具和行政工具的划分方法。供给型、环境型和需求型政策工具分类法在既有研究中最为常见，该方法综合考虑了政府干预手段和干预程度。

④“北大法宝在线数据库”网址为https://libyw.ucas.ac.cn/https/1SjapK4g9QWMKZdqGX5SBIacUGddK7Z07CpDY/.

表1　部分数据主权政策文本

Table 1　Part of data sovereignty policy text

序号	政策名称	效力级别	发布部门	发布日期	实施日期
1	《中华人民共和国个人信息保护法》	法律	全国人大常委会	2021年8月20日	2021年11月1日
2	《中华人民共和国海南自由贸易港法》	法律	全国人大常委会	2021年6月10日	2021年6月10日
3	《中华人民共和国数据安全法》	法律	全国人大常委会	2021年6月10日	2021年9月1日
4	《中华人民共和国网络安全法》	法律	全国人大常委会	2016年11月7日	2017年6月1日
5	《国务院关于印发"十四五"数字经济发展规划的通知》	国务院规范性文件	国务院	2021年12月12日	2021年12月12日
6	《国务院关于印发促进大数据发展行动纲要的通知》	国务院规范性文件	国务院	2015年8月31日	2015年8月31日
7	《数据出境安全评估办法》	部门规章	国家互联网信息办公室	2022年7月7日	2022年9月1日

表2　部分数据主权政策文本内容编码情况

Table 2　Encoding of text content of some data sovereignty policies

序号	政策名称	政策分析单元	编码	政策工具
1	《中华人民共和国个人信息保护法》	第十条　任何组织、个人不得非法收集、使用、加工、传输他人个人信息，不得非法买卖、提供或者公开他人个人信息；不得从事危害国家安全、公共利益的个人信息处理活动。	1-1	法规管制
		……		
4	《中华人民共和国数据安全法》	第十一条　国家积极开展数据安全治理、数据开发利用等领域的国际交流与合作，参与数据安全相关国际规则和标准的制定，促进数据跨境安全、自由流动。	4-9	国际合作
		……		
22	《数据出境安全评估办法》	第三条　数据出境安全评估坚持事前评估和持续监督相结合、风险自评估与安全评估相结合，防范数据出境安全风险，保障数据依法有序自由流动。	22-1	审查评估

出现在2015年国务院印发的《促进大数据发展行动纲要》，指出应"增强网络空间数据主权保护能力，维护国家安全，有效提升国家竞争力"。然而，早在2010年已经开始出现互联网主权的概念。数据主权政策发布数量在2016年和2021年迎来了2次大幅提升(图2)。结合大数据、云计算和区块链等新技术的出现与快速发展，将数据主权政策演化划分为4个阶段。

(1) 互联网主权阶段（2010—2013年）。2010年，国务院新闻办公室发布《中国互联网状况》白皮书，指出"中华人民共和国境内的互联网属于中国主权管辖范围，中国的互联网主权应受到尊重和维护"。中国逐步完善互联网法律法规，积极推动互联网基础设施建设，加强对互联网的管理和监管，确保互联网安全，推动互联网产业发展。

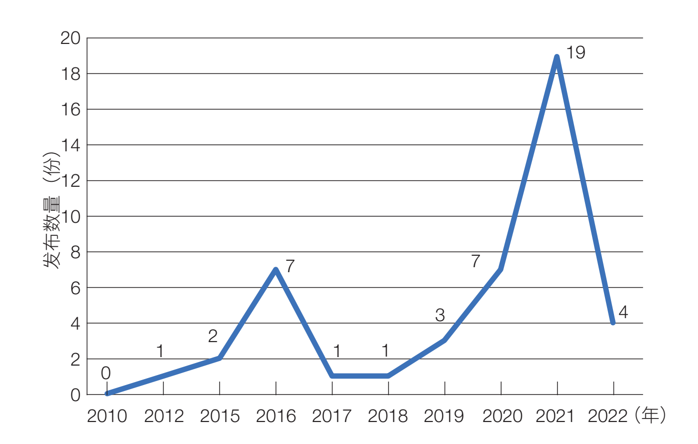


图2　2010—2022年我国数据主权政策发文时间分布

Figure 2　Time distribution of China's data sovereignty policy from 2010 to 2022

（2）信息主权阶段（2014—2015年）。2014年7月16日，习近平主席在巴西国会发表题为《弘扬传统友好 共谱合作新篇》演讲，强调每一个国家在信息领域的主权权益都不应受到侵犯。这一阶段，相关政策强调加强信息自主可控，建设信息基础设施，推进信息安全技术和标准的研发，提升中国在信息主权领域的实际管控能力和国际话语权。

（3）网络空间主权阶段（2015—2018年）。2015年7月颁布实施的《中华人民共和国国家安全法》首次明确了网络空间主权的概念。在2017年3月发布的《网络空间国际合作战略》中将“坚定维护中国网络主权、安全和发展利益”作为中国参与网络空间国际合作的战略目标之一。2018年5月，在国家互联网信息办公室发布的《数字中国建设发展报告（2017年）》中，将“维护网络主权”作为数字中国建设面临的形势和努力方向。这一阶段，相关政策强调加强网络安全保护、建立网络主权管理机制、推进网络空间治理等，以确保国家在网络空间中有权利和能力保护国家安全和维护国家利益。

（4）数据主权阶段（2019年至今）。2019年7月，《加强工业互联网安全工作的指导意见》指出，依据工业门类领域、数据类型、数据价值等建立工业互联网数据分级分类管理制度，开展重要数据出境安全评估和监测。2020年9月，中国发起《全球数据安全倡议》，呼吁各国未经他国法律允许不得直接向企业或个人调取位于他国的数据。2021年9月开始施行的《中华人民共和国数据安全法》规定，维护数据安全，应当坚持总体国家安全观，建立健全数据安全治理体系，提高数据安全保障能力；同年，《中华人民共和国个人信息保护法》也对个人信息的跨境流动、数据本地化存储及域外效力做出了相关规定。这一阶段，相关政策强调建立数据分级分类制度、明确数据出境安全评估要求、推进数据保护能力认证机制等，以确保数据在跨境流动过程中得到有效保护和合规使用。

综上，中国的数据主权战略经历了“互联网主权—信息主权—网络空间主权—数据主权”的演化，整体政策特点为以安全为纲，推动定规立制，促进数据有序流动，加强数据跨境流动过程和出境后的风险评估和监管。

2.2 政策发文机构分布

政策发文机构能够反映出各政策法规的效力级别。国务院、工业和信息化部、商务部、全国人大常委会等共计33个机构参与数据主权政策制定，表3展示了发布政策的核心机构分布情况。我国数据主权政策有联合发布和单独发布2种形式；45份中央法规政策文件中，8份为联合发布，37份为单独发布，单独发布占比82.2%。

基于统计结果可知，发布政策文件较多的机构依次为国务院、工业和信息化部、商务部；联合发文的主要为国家发展和改革委员会、工业和信息化部。政策发布核心主体[⑤]共有5个，占比15.2%，非核心主体共有28个，占比84.8%，各政策主体在整体分布上较为分散。总体来看，参与数据主权政策制定的主体众多，但核心主体的发文数量占比54.2%，较为集中。整体呈现出以国务院、工业和信息化部、商务部、全国人大常委会、国家互联网信息办公室为五大中心发文的结构特征（表3）。

2.3 政策类型分布

政策发布所采用的文种类型因政策性质和目标的不同而有所差异，由此导致了不同政策的效力不同[13,14]。本文所收集的政策样本中，数据主权政策采用了9种不同的形式进行发布，包括方案、意见、规划、法律、通知、纲要和办法等，展现出形式多样性（表4）。其中，方案、意见和规划在政策文本中占比

⑤ 发布数量超过3份的部门被视为核心主体。

表3 2010—2022年我国数据主权政策核心发布机构分布情况
Table 3 Distribution of government agencies issuing China's data sovereignty policies from 2010 to 2022

发布机构	发布数量(份)
国务院	14
工业和信息化部	10
商务部	8
全国人大常委会	4
国家互联网信息办公室	3

较高，这说明现有政策中相关部门提出的见解和处理意见较多；而针对性更强、更具体的办法在整体政策分布中较少；全国人大及其常务委员会共颁布4项法律。

3 数据主权政策二维分析

政策工具是政策主体为实现政策目标而采取的具体措施和手段，本文根据已有研究[15-17]从需求型、环境型和供给型3个方面构建了数据主权政策工具分析维度，通过政策工具识别出政策实施的重要途径和保障措施，并基于LDA主题模型进行政策主题分析，揭示了政策的核心内容和主要目标，以期为国家进一步优化完善数据主权政策制定提供决策参考。

表4 2010—2022年我国数据主权政策文种类型分布
Table 4 Type distribution of China's data sovereignty policy documents from 2010 to 2022

政策类型	发布数量(份)	百分比
方案	11	24.4%
意见	10	22.2%
规划	7	15.6%
法律	4	8.9%
通知	4	8.9%
纲要	3	6.7%
报告	3	6.7%
战略	2	4.4%
办法	1	2.2%

考虑到一个政策分析单元可能同时应用多种政策工具，因此本文用于分析的政策工具数量之和超过编码单元总数。从政策工具应用类型来看，我国数据主权政策对政策工具的应用较为全面，需求型、环境型、供给型政策工具分别占比21.6%、59.6%、18.8%（图3）。其中，环境型政策工具较为常用，需求型和供给型政策工具应用频率次之。我国采用多种政策工具共同推动数据跨境安全流动和数据主权保护。

3.1 政策工具维度的实证分析

（1）需求型政策工具，旨在引入各方力量进行交流与合作，积极开展数据安全管理、跨境流动和国际合作规则等试点，突出重要领域和关键环节，从而促进数据安全有序流动（表5）。需求型数据主权政策工具以试点示范和国际合作为主，着重于通过试点示范和国际合作等发展模式先行先试，积累数据安全管理和数据主权保护有效经验并推广普及于全社会，注重引导社会力量参与，开展宣传推广，从而推动行业规范发展、提高数据主权保护意识。

（2）环境型政策工具，旨在通过优化各主体、各环节实施的环境条件、政策导向、标准规范和问责机制等引导和监管各主体采取更积极的数据安全管理和数据主权保护措施（表6）。整体而言，环境型政策工

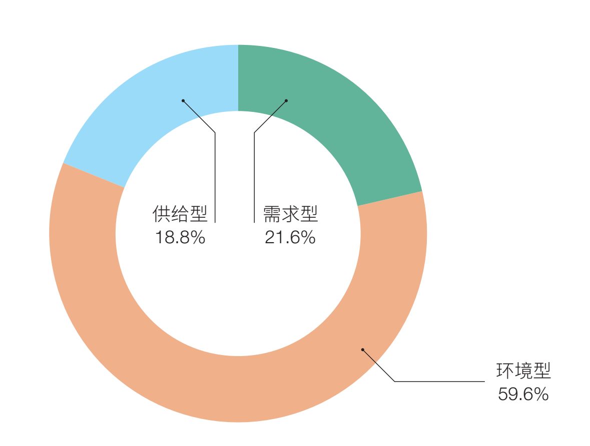


图3 2010—2022年我国数据主权政策工具频率分布
Figure 3 Frequency distribution of China's data sovereignty policy instruments from 2010 to 2022

表5　需求型数据主权政策工具的内涵与维度

Table 5　Connotation and dimensions of demand-based data sovereignty policy instruments

工具名称	具体内涵	关键词
试点示范	开展数据安全管理、跨境流动和国际合作规则等试点，突出重要领域和关键环节，及时总结推广地方经验，发挥重点改革的突破性作用和试点的带动效应	试点、试验
服务外包[18]	政府鼓励或组织社会第三方机构参与到数据安全评估、数据保护认证等过程	第三方
社会参与	引导社会各方力量共同参与数据共享、数据安全、数据主权保护等工作，实现多方融合、高效协作，如鼓励政企、校企合作	调动、激励
宣传推广	支持开展数据安全知识宣传普及，加大对数据管理、数据开放共享等国家标准的宣贯力度，合理引导社会舆论，营造良好氛围	宣传、教育
国际合作	积极参与国际规则制定，促进跨境数据共享和技术合作，加强保护治理合作等	国际交流、合作

具占比超过半数，作为一种间接调控手段[19]受到政策主体的青睐。法规管制是常用方法，随着新技术的发展，数据规模不断扩大，国际数据主权威胁等问题日益凸显，需要加强规制积极应对域外效力等问题。标准规范、安全规范、保障措施和审查评估是引导新兴数字产业规范发展、保障数据安全有序流动等正在逐步完善的政策文件，包括跨境流通准则、安全评估和保障机制等。策略性措施对政策体系起到补充作用。当前知识产权政策工具应用较少，是数据主权政策中的明显空白点。

（3）供给型政策工具，旨在为各主体的数据安全管理、数据主权保护、信息基础设施建设、技术研发、人才培养、统筹协调等提供各类资源支持（表7）。在供给型政策工具中，信息基础设施建设、技术支持和组织建设占比较高。随着大数据、云计算和区块链等新技术的出现，信息化设施成为数字经济发展的重要基础条件。《中华人民共和国国民经济和社会发展第十四个五年规划和2035年远景目标纲要》和《中华人民共和国网络安全法》均提出要建立健全关键信息基础设施保护体系。技术支持是保障数据安全

表6　环境型数据主权政策工具的内涵与维度

Table 6　Connotation and dimensions of environmental data sovereignty policy instruments

工具名称	具体内涵	关键词
法规管制	政府通过设定法律法规、部门规章、工作办法、问责机制等强制措施来加强对数据主权、数据安全、数据跨境等的监管与保护	法规、监管
审查评估	政府加强对数据安全的审查，并对各地试点工作开展情况进行评估	审查、评估
标准规范	政府制定与数据主权、数据安全、数据共享等相关政策的标准规范	标准、规范、规则
目标规划	加强数据资源规划，对数据安全、数据主权做出总体目标和规划	目标、规划
策略性措施	政府为加强数据安全、数据共享、数据跨境、数据主权保护等所制定的各项策略性措施，如鼓励创新、鼓励协同融合、统筹兼顾、探索培育新业态等	创新、探索、统筹、鼓励
保障措施	为保障数据安全、数据主权保护、数据跨境流动等工作，制定的各项保障措施，如推进协作、强化监管、机制建设等	机制、保障
平台建设	政府搭建数据开放、共享、交易、监管、预测平台	平台
安全规范	政府制定数据安全相关工作的安全标准、行业规范等	安全、规范
知识产权	强化数据及知识产权保护，提高知识产权保护意识	知识产权

流动的重要力量，应通过开发核心技术构建安全可靠的数据流通环境。在组织建设的统筹协调下，信息基础设施建设和技术支持相辅相成，为数据安全流动和数据主权保护提供持续动力。当前人才支持和资金支持政策工具应用较少，分别为6.67%和8.89%。资金支持具有定向性优势，人才培养则具有长期效益，应适当增加这两个政策工具的应用。

3.2 政策主题维度的实证分析

3.2.1 LDA主题模型主题数与一致性分析

LDA主题模型可用于挖掘大量文本的潜在主题特征。它是一种无监督的、非结构化的概率模型，运用了单词共现具有语义关联的规律，无需预先设置字典或主题类别便可发现文档语料库中的潜在主题[20]，减少了研究者主观判断对主题分类的影响[21]。本文将每个编码单元视为一个文档，选择主题一致性指标模型用于判定最优主题数量，若一致性处于较高水平，模型的主题结构将更加稳定。本文选用“哈工大停用词表”⑥对文档进行文本预处理（中文分词、去停用词），主题数与一致性结果如图4所示。通过主题评估，本文选用3作为主题个数，并使用pyLDAvis算法对LDA主题模型结果进行可视化分析与展示。

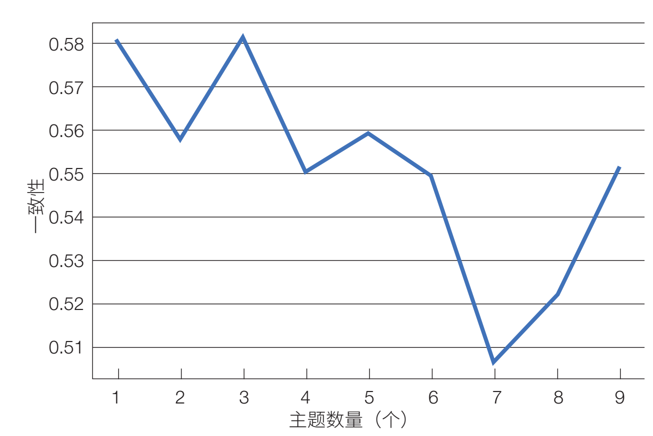


图4 2010—2022年我国数据主权政策主题数划分的一致性分析

Figure 4 Coherence analysis of division of topic numbers in China’s data sovereignty policy from 2010 to 2022

3.2.2 词云可视化

图5展示了基于LDA主题模型挖掘出的数据主权政策的主题词袋（bag of words）⑦。根据LDA主题模型可视化结果，当前数据主权政策分为3类：数据安全与个人信息保护；数据跨境与国际合作；数据安全评估与数据出境。

（1）数据安全与个人信息保护（图5a）。推动建设全面数据安全保护体系，包括完善数据分级分类制度和个人信息授权使用制度。为形成数据资源汇集共

表7 供给型数据主权政策工具的内涵与维度

Table 7 Connotation and dimensions of supply-based data sovereignty policy instruments

工具名称	具体内涵	关键词
信息基础设施建设	加强与数据主权、数据跨境、数据安全相关的信息或数字基础设施的建设与提升	信息基础设施保护、数字基础设施建设
人才支持	加大对技术研发、数据安全和涉外法治等人才的培养，培养复合型国际人才	人才培养、教育培训
资金支持	政府层面对数据主权相关政策的制定与实施给予财政补助和资金支持	财政投入、资金支持
政策支持	加大对跨境贸易、数据分类、数据安全规范等方面的政策支持力度	政策支持
技术支持	加强数据安全、网络安全等关键技术的研发、创新与应用	技术研发、技术创新
组织建设	加强数据主权保护、数据跨境、数据安全等工作的组织领导和统筹并协调各部门分工，协调监管与保护，切实加强工作指导	组织领导、统筹协调、监管协作

⑥ 哈工大停用词表是由哈尔滨工业大学自然语言处理实验室发布的一个停用词表。它包含了常用的中文停用词，如“的”“是”“在”“你们”等。该停用词表的规模较大，可用于大多数中文自然语言处理任务。停用词是在文本处理中被过滤或忽略的常见词汇，它们在文本中出现的频率较高，但往往对文本含义和主题没有实质性贡献。

⑦ 在一篇文档中仅考虑词汇是否出现而不考虑出现的顺序。

享、数据流动安全有序的数据要素良性发展格局，应妥善处理国家安全、数据跨境流通和个人隐私保护三者的关系。

（2）数据跨境与国际合作（图5b）。明确数据安全法在域外的适用效力，进一步健全数据跨境流通规范，实施数据跨境传输与安全管控试点。探索加入区域性国际数据跨境流通规则制定，促进形成数据跨境流通的全球协同机制，强化中国同世界其他国家和地区之间的安全协作与信息资源共享。

（3）数据安全评估与数据出境（图5c）。健全数据分级分类管理制度和数据出境安全审查机制。确保数据在出境前能够进行全面的事前评估、持续监督和风险自评估，有效识别和防范数据出境带来的安全风险，保护国家和个人的敏感信息。探索数据保护能力认证机制设计，为数据出境提供客观和可信的评估标准，确保数据的合法性、安全性和可控性。

综上，中国数据主权政策主要运用了环境型政策工具，辅之需求型和供给型政策工具，推动数据跨境安全流动和数据主权保护。政策文本内容呈现出“数据安全与个人信息保护”“数据跨境与国际合作”“数据安全评估与数据出境”三大主题特征。我国陆续制定了《中华人民共和国个人信息保护法》《中华人民共和国数据安全法》等法律，出台了《数据出境安全评估办法》等一系列规章制度，实行严格数据出境和信息保护管理模式，确保我国数据主权得以保护。

4 启示与建议

本文在对数据主权领域政策进行分析的基础上，对我国数据跨境安全流动与数据主权保护提出4条政策建议。

（1）积极主导和参与国际规则制定，建立互信机制，提高国际话语权。积极主导和建设性参与制定、完善数字时代的国际规则和标准，与更广泛的国家建立多种形式的数据合作平台，就数据基础设施建设、数据资源开发利用、数据安全保障等方面进行技术支持和能力建设，促进数据互联互通和共享共赢。推动与其他国家或地区在数据保护水平、标准、规范等方面的合作讨论，争取更多一致意见和共识，实现数据

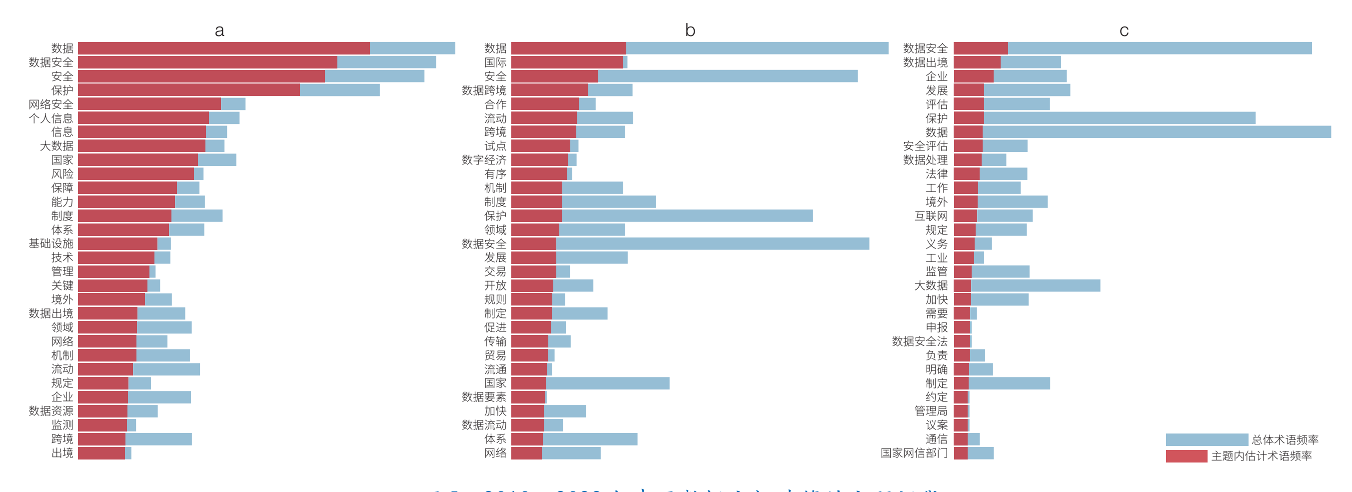


图5　2010—2022年我国数据主权政策的主题词袋

Figure 5　Bag of words for China’s data sovereignty policy from 2010 to 2022

（a）主题1“数据安全与个人信息保护”；（b）主题2“数据跨境与国际合作”；（c）主题3“数据安全评估与数据出境”；图中为各主题按出现概率由高到低排序的前30个词，代表了各自对主题的贡献度

(a) Topic 1 “data security and personal information protection”; (b) Topic 2 “data cross-border and international cooperation”; (c) Topic 3 “data security assessment and data export”; The figure shows top 30 most relevant terms for topics, representing their contributions to the topic

保护标准和规范的相互认可或互惠适当性。

（2）优化数据出境安全评估流程，提高评估效率和准确性。建立以风险为导向的数据分类管理制度，对不同类型和等级的数据采取差异化的管控措施，加强数据安全评估的标准制定和实施，明确数据安全评估的要求和指导，确保数据出境过程中的安全性和可信性。利用先进技术手段提高数据出境安全评估的效率和准确性。例如，可以通过引入人工智能和大数据分析技术等自动化评估流程，快速识别高风险的数据出境行为，以提高评估结果的准确性和可靠性。

（3）完善个人信息出境标准合同模版，提高合规运营的效率。《个人信息出境标准合同办法》于2023年6月1日起施行，旨在确保个人信息合法、安全、有序地跨境传输。通过使用标准合同模板，我国能够通过《个人信息出境标准合同办法》赋予的法律约束力来将境内的管辖权“延伸”至境外，达到一定“境内法域外适用”的效果，实现数据跨境流动保护。未来，可以将标准合同模板进行模块化拓展，根据组织、企业或个人的业务需求，制定更多可以选择的模块，以降低相关实体合规运营成本，提高效率。

（4）强化数据安全法治保障，构筑数据主权防线。以总体国家安全观为指导，完善保障数据安全的法律法规，强化数据安全治理体系。明确数据安全的法律责任，保护关键数据基础设施；建立数据安全风险评估和应急响应机制；开展数据主权与安全宣传教育，提高全民数据安全意识和能力；推动国际合作，加强与其他国家和地区在数据安全法律法规方面的交流和互鉴，共同促进国际数据安全标准制定，提升中国在全球数据治理中的话语权和影响力。

参考文献

1 洪永淼, 张明, 刘颖. 推动跨境数据安全有序流动 引领数字经济全球化发展. 中国科学院院刊, 2022, 37(10): 1418-1425.

Hong Y M, Zhang M, Liu Y. Promoting safe and orderly flow of cross-border data to lead development of globalization of digital economy. Bulletin of Chinese Academy of Sciences, 2022, 37(10): 1418-1425. (in Chinese)

2 冉从敬, 刘妍. 数据主权的理论谱系. 武汉大学学报(哲学社会科学版), 2022, 75(6): 19-29.

Ran C J, Liu Y. Outline of data sovereignty genealogies. Wuhan University Journal (Philosophy & Social Science), 2022, 75(6): 19-29. (in Chinese)

3 何波. 数据主权的发展、挑战与应对. 网络信息法学研究, 2019, (1): 201-216.

He B. The development, challenges, and countermeasures of data sovereignty. Journal of Cyber and Information Law, 2019, (1): 201-216. (in Chinese)

4 Rainie S C, Schultz J L, Briggs E, et al. Data as a strategic resource: Self-determination, governance, and the data challenge for indigenous nations in the United States. International Indigenous Policy Journal, 2017, 8(2), doi: 10.18584/iipj.2017.8.2.1.

5 Hippelainen L, Oliver I, Lal S. Towards dependably detecting geolocation of cloud servers// International Conference on Network and System Security. Cham: Springer, 2017: 643-656.

6 孙南翔, 张晓君. 论数据主权——基于虚拟空间博弈与合作的考察. 太平洋学报, 2015, 23(2): 63-71.

Sun N X, Zhang X J. On data security—An examination of game and cooperation in cyberspace. Pacific Journal, 2015, 23(2): 63-71. (in Chinese)

7 齐爱民, 盘佳. 数据权、数据主权的确立与大数据保护的基本原则. 苏州大学学报(哲学社会科学版), 2015, 36(1): 64-70.

Qi A M, Pan J. Right to data, data sovereignty and the basic principle of big data protection. Journal of Soochow University (Philosophy & Social Science Edition), 2015, 36(1): 64-70. (in Chinese)

8 Jarke M, Otto B, Ram S. Data sovereignty and data space ecosystems. Business & Information Systems Engineering, 2019, 61(5): 549-550.

9 Alboaie S, Cosovan D. Private data system enabling self-sovereign storage managed by executable choreographies//

IFIP International Conference on Distributed Applications and Interoperable Systems. Cham: Springer, 2017: 83-98.

10 冉从敬. 数据主权治理的全球态势与中国应对. 人民论坛, 2022, (4): 24-27.

Ran C J. Global situation of data sovereignty governance and China's response. People's Tribune, 2022, (4): 24-27. (in Chinese)

11 郑琳, 李妍, 王延飞. 新时代国家数据主权战略研究. 情报理论与实践, 2022, 45(6): 55-60.

Zheng L, Li Y, Wang Y F. Research on national data sovereignty strategies in the new era. Information Studies (Theory & Application), 2022, 45(6): 55-60. (in Chinese)

12 吕文晶, 陈劲, 刘进. 政策工具视角的中国人工智能产业政策量化分析. 科学学研究, 2019, 37(10): 1765-1774.

Lv W J, Chen J, Liu J. A quantitative analysis of China's Artificial Intelligence industry policy from the perspective of policy tools. Studies in Science of Science, 2019, 37(10): 1765-1774. (in Chinese)

13 陈玲, 段尧清. 我国政府开放数据政策的实施现状和特点研究:基于政府公报文本的量化分析. 情报学报, 2020, 39(7): 698-709.

Chen L, Duan Y Q. Analyzing implementation of the Chinese government open data policy using government bulletin text as example. Journal of the China Society for Scientific and Technical Information, 2020, 39(7): 698-709. (in Chinese)

14 黄如花, 温芳芳. 我国政府数据开放共享的政策框架与内容:国家层面政策文本的内容分析. 图书情报工作, 2017, 61(20): 12-25.

Huang R H, Wen F F. Policy framework and content of opening and sharing government data in China: A content analysis of policy documents at the national level. Library and Information Service, 2017, 61(20): 12-25. (in Chinese)

15 Rothwell R. Reindustrialization and technology: Towards a national policy framework. Science and Public Policy, 1985, 12(3): 113-130.

16 黄萃, 徐磊, 钟笑天, 等. 基于政策工具的政策-技术路线图(P-TRM)框架构建与实证分析——以中国风机制造业和光伏产业为例. 中国软科学, 2014, (5): 76-84.

Huang C, Xu L, Zhong X T, et al. A framework of integrated policy-technology roadmap (P-TRM) and its use: Examples of wind turbine and solar PV industries. China Soft Science, 2014, (5): 76-84. (in Chinese)

17 付琳, 张东雨, 闫昊本, 等. 基于政策文本分析的中国碳减排政策工具研究. 科学学研究, 2023, 41(3): 435-443.

Fu L, Zhang D Y, Yan H B, et al. Research on China's policy instruments of carbon emission mitigation through policy content analysis method. Studies in Science of Science, 2023, 41(3): 435-443. (in Chinese)

18 李霞, 陈琦, 贾宏曼. 中国智慧城市政策体系演化研究. 科研管理, 2022, 43(7): 1-10.

Li X, Chen Q, Jia H M. Research on evolution of the smart city policy system in China. Science Research Management, 2022, 43(7): 1-10. (in Chinese)

19 时颖惠, 薛翔. 政策工具视角下我国信息安全政策研究——基于81份政策文本的量化分析. 现代情报, 2022, 42(1): 130-138.

Shi Y H, Xue X. Research on information security policy in China from the perspective of policy tools—Based on the quantitative analysis of 81 policy texts. Journal of Modern Information, 2022, 42(1): 130-138. (in Chinese)

20 张潮, 冷伏海. 机器学习辅助智能决策分析——以"绿色创新"问题为例. 科研管理, 2022, 43(9): 32-40.

Zhang C, Leng F H. An analysis of the machine-learning-assisted intelligent decision-making—A study by taking "green innovation" as an example. Science Research Management, 2022, 43(9): 32-40. (in Chinese)

21 Huang A H, Lehavy R, Zang A Y, et al. Analyst information discovery and interpretation roles: A topic modeling approach. Management Science, 2018, 64(6): 2833-2855.

Research on Chinese data sovereignty policy based on LDA model and policy instruments

QIAO Han* XU Junru

（1 School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;

2 MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation,

University of Chinese Academy of Sciences, Beijing 100190, China）

Abstract Data sovereignty has become an important component of national sovereignty in the dual context of the digital economy development and the overall national security concept. Major countries and regions are actively carrying out data sovereignty strategic deployment and engaging in fierce competition in data resources, data technology, and data rules. This work adopts the policy text analysis method to study China's data sovereignty policy, and employs the LDA model and policy instruments to quantitatively analyze the process evolution and thematic characteristics of China's data sovereignty policy. Drawing on these findings, this study comprehensively considers the global data sovereignty policy and puts forward four policy recommendations: actively lead and participate in the formulation of international rules, optimize the data export security assessment process, improve the standard contract template for personal information export, and strengthen the legal protection of data security.

Keywords data sovereignty, data security, policy instruments, Latent Dirichlet Allocation (LDA)

乔 晗 中国科学院大学经济与管理学院教授，博士生导师，《管理评论》常务副主编。研究领域为数字经济、科技创新与商业模式创新、风险投资与创业等。E-mail: qiaohan@ucas.ac.cn

QIAO Han Professor and Doctoral Supervisor of School of Economics and Management, University of Chinese Academy of Sciences (UCAS); Executive Associate Editor-in-Chief of *Management Review*. Her main research areas cover digital economy, scientific and technological innovation and business model innovation, venture capital and entrepreneurship, etc.
E-mail: qiaohan@ucas.ac.cn

■责任编辑：岳凌生

*Corresponding author